DOCUMENT RESUME

ED 201 657 TM 810 249

AUTHOR Malgady, Robert G.
TITLE Note on Contrasting Part Correlations in Regression Models.
PUB DATE [75]
NOTE 13p.
AVAILABLE FROM Robert G. Malgady, Department of Educational Statistics, 23 Press Building, New York University, New York, NY 10003.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Correlation; *Mathematical Models; *Multiple Regression Analysis; *Predictor Variables; Statistical Significance; *Suppressor Variables
IDENTIFIERS *Contrasting Groups Method; Hotellings t

ABSTRACT

Common applications of the part correlation coefficient are in causal regression models and estimation of suppressor variable effects. However, there is no statistical test of the significance of the difference between a zero-order correlation and a part correlation, nor between a pair of part correlations. Hotelling's t is used for contrasting: (1) a zero-order predictor-criterion correlation, and a part correlation when the influence of a covariate is removed from predictor variance; (2) two part correlations based on mutual predictors and covariates; (3) two part correlations based on the same predictor but different covariates; and (4) two part correlations based on different predictors and the same covariate. Although Hotelling's t has enjoyed widespread citation in most research literature, criticisms have been raised because of its assumption of conditional normality and because, strictly speaking, statistical inference is limited to the sample values of the variables. However, Hotelling's t has greater familiarity and ubiquity, and its assumptions are no more restrictive than those associated with the Pearson product-moment correlation. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED201657

NOTE ON CONTRASTING PART CORRELATIONS IN REGRESSION MODELS

Robert G. Malgady

New York University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. G. Malgady

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Part Correlation

Date Submitted:

Author's Address: Department of Educational Statistics
23 Press Building
New York University
New York, New York 10003

TM 810 299

*Requests for reprints should be addressed to Robert G. Malgady, Department of Educational Statistics, 23 Press Building, New York University, New York, New York 10003.

Abstract

Common applications of the part correlation coefficient are in causal regression models and estimation of suppressor variable effects. However, there is no statistical test of the significance of the difference between a zero-order correlation and a part correlation, nor between a pair of part correlations. This note uses Hotelling's (1940) *t* for contrasting dependent *rs* for contrasting a zero-order and first order part correlation, and for contrasting first-order part correlations based on different predictors and different covariates, the same predictor but different covariates, and different predictors and the same covariate.

NOTE ON CONTRASTING PART CORRELATIONS IN REGRESSION MODELS

The part or semipartial correlation coefficient is useful for estimating the contribution of a variable to prediction of some criterion, while controlling extraneous variance in either the predictor or criterion. Part correlation is an important statistic in the development of causal regression models, such as in path analysis (e.g., Spaeth, 1975); moreover, it is the basis for definition and estimation of suppressor variable effects (Conger & Jackson, 1972), and of a variable's unique contribution to criterion prediction in hierarchical multiple regression (Cohen & Cohen, 1975). Unfortunately, there is no statistical test for determining whether part correlation is "effective" relative to zero-order correlation, nor for that matter, for making pairwise contrasts among different part correlations appearing in causal or hierarchical regression models. Consequently, the purpose of the present note is to suggest a statistical test of the difference between (1) a zero-order predictor-criterion correlation, r_{ij}, and a part correlation when the influence of a covariate is removed from predictor variance, $r_{i(j.k)}$; (2) two part correlations based on mutual predictors and covariates, $r_{i(j.k)}$ versus $r_{i(k.j)}$; (3) two part correlations based on the same predictor but different covariates, $r_{i(j.k)}$ versus $r_{i(j.m)}$; and (4) two part correlations based on different predictors and the same covariate, $r_{i(j.m)}$ versus $r_{i(k.m)}$.

Thus, consider the general situation in which a group of N individuals are measured on a criterion variable (X_i) and a predictor variable (X_j), where the researcher wishes to eliminate variance in X_j due to a third variable or covariate (X_k). Further, consider a second transformed predictor variable representing that portion of X_j that is linearly independent of

X_k (say X_{j^*}). Then using Hotelling's (1940) well known t to compare the two predictors X_j and X_{j^*} of the criterion X_i yields:

$$t = \frac{(r_{ij} - r_{ij^*})[(N - 3)(1 + r_{jj^*})]^{1/2}}{[2(1 - r_{ij}^2 - r_{ij^*}^2 - r_{jj^*}^2 + 2r_{ij}\, r_{ij^*}\, r_{jj^*})]^{1/2}}. \quad (1)$$

But since $r_{ij^*} = r_{i(j.k)}$ and $r_{jj^*} = (1 - r_{jk}^2)^{1/2}$, Equation (1) can be expressed in terms of the correlations of interest as:

$$t = \frac{(r_{ij} - r_{i(j.k)})[(N - 3)(1 + (1 - r_{jk}^2)^{1/2})^{1/2}]}{[2(r_{jk}^2 - r_{ij}^2 - r_{i(j.k)}^2 + 2r_{ij}\, r_{i(j.k)}(1 - r_{jk}^2)^{1/2})^{1/2}]}, \quad (2)$$

which therefore tests the null hypothesis that $\rho_{ij} = \rho_{i(j.k)}$ with $N - 3$ degrees of freedom (df).

In the second case, for determining whether two different predictors explain different amounts of criterion variance while controlling their mutual influence on each other, consider the additional predictor variable representing that portion of X_k that is linearly independent of X_j (say X_{k^*}). Then Hotelling's t to compare the two predictors X_{j^*} and X_{k^*} of the criterion X_i yields:

$$t = \frac{(r_{ij^*} - r_{ik^*})[(N - 3)(1 + r_{j^*k^*})]^{1/2}}{[2(1 - r_{ij^*}^2 - r_{ik^*}^2 - r_{j^*k^*}^2 + 2r_{ij^*}\, r_{ik^*}\, r_{j^*k^*})]^{1/2}}. \quad (3)$$

Clearly, $r_{ik^*} = r_{i(k.j)}$ and it can be shown that $r_{j^*k^*} = (r_{jk}^3 - r_{jk}^4)/(1 - r_{jk}^2)$ (see Appendix A). Thus, Equation (3) can be expressed as:

$$t = \frac{(r_{i(j.k)} - r_{i(k.j)})\left[(N-3)\left(1 + \frac{r_{jk}^3 - r_{jk}}{1 - r_{jk}^2}\right)\right]^{1/2}}{\left[2(1 - r_{i(j.k)}^2 - r_{i(k.j)}^2 - \frac{(r_{jk}^3 - r_{jk}^2)}{(1 - r_{jk}^2)^2} + 2r_{i(j.k)}\, r_{i(k.j)} \frac{r_{jk}^3 - r_{jk}}{1 - r_{jk}^2}\right)\right]^{1/2}} \quad (4)$$

which tests the null hypothesis that $\rho_{i(j.k)} = \rho_{i(k.j)}$ with df = $N - 3$.

In the third case, for determining whether two different covariates influence a predictor-criterion correlation differently, consider the transformed predictor variable representing that portion of X_j that is linearly independent of still a fourth variable X_m (say $X_{j^{**}}$). Then Hotelling's formula to compare the two predictors X_{j^*} and $X_{j^{**}}$ of the criterion X_i yields:

$$t = \frac{(r_{ij^*} - r_{ij^{**}})\left[(N-3)(1 + r_{j^*j^{**}})\right]^{1/2}}{\left[2(1 - r_{ij^*}^2 - r_{ij^{**}}^2 - r_{j^*j^{**}}^2 + 2r_{ij^*}\, r_{ij^{**}}\, r_{j^*j^{**}})\right]^{1/2}} \quad (5)$$

It follows that $r_{ij^{**}} = r_{i(j.m)}$ and it can be shown (see Appendix B) that

$$r_{j*j**} = \frac{1 - r_{jm}^2 - r_{jk}^2 + r_{jk}\, r_{km}\, r_{jm}}{(1 - r_{jk}^2)^{1/2}\,(1 - r_{jm}^2)^{1/2}}.$$

Thus, in this case Equation (5) can be expressed as:

Insert Equation (6) about here

---

which tests the null hypothesis that $\rho_{i(j.k)} = \rho_{i(j.m)}$ with df = $N - 4$.

Finally, in the fourth case, for comparing the effect of the same covariate on two different predictor-criterion correlations, consider the transformed predictor variable representing that portion of X_k that is linearly independent of X_m (say X_{k**}). Hotelling's contrast of the correlations of predictors X_{j**} and X_{k**} with criterion X_i yields:

$$t = \frac{(r_{ij**} - r_{ik**})\,[(N - 3)(1 + r_{j**k**})]^{1/2}}{[2(1 - r_{ij**}^2 - r_{ik**}^2 - r_{j**k**}^2 + 2r_{ij**}\, r_{ik**}\, r_{j**k**})]^{1/2}}. \qquad (7)$$

where $r_{ik**} = r_{i(k.m)}$ and $r_{j**k**} = r_{jk.m}$. Thus, Equation (7) can be expressed as:

$$t = \frac{(r_{i(j.m)} - r_{i(k.m)})\,[(N - 4)(1 + r_{jk.m})]^{1/2}}{[2(1 - r_{i(j.m)}^2 - r_{i(k.m)}^2 - r_{jk.m}^2 + 2r_{i(j.m)}\, r_{i(k.m)}\, r_{jk.m})]^{1/2}}, \qquad (8)$$

which tests the null hypothesis that $\rho_{i(j.m)} = \rho_{i(k.m)}$ with df = $N - 4$.

Discussion

Although Hotelling's t has enjoyed widespread citation in most statistical textbooks and the research literature, criticisms have been raised because of its assumption of conditional normality and because, strictly speaking, statistical inference is limited to the sample values of the variables. Williams (1959) has published an alternative to Hotelling's t which relaxes the conditional normality assumption; and Olkin (1967) has proposed instead a z statistic which permits unrestricted generalization to the universe of each variable's values, but which assumes multivariate normality. Hendrickson, Stanley, and Hills (1970) have shown, however, that there is little systematic difference in outcome when the three competing statistics are applied to the same data, particularly when N is large, when the difference between correlations with the criterion is small, or when the two predictors are highly correlated. Thus, the present developments were cast in terms of Hotelling's t because of its greater familiarity and ubiquity, and also because its assumptions are no more restrictive than those associated with the Pearson product-moment correlation itself.

References

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N.J.: Erlbaum, 1975.

Conger, A.J., & Jackson, D.N. Suppressor variables, prediction, and the interpretation of psychological relationships. Educational and Psychological Measurement, 1972, 32, 579-599.

Hendrickson, G.F., Stanley, J.C., & Hills, J.R. Olkin's new formula for significance of r_{13} vs. r_{23} compared with Hotelling's method. American Educational Research Journal, 1970, 7, 189-195.

Hotelling, H. The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Annals of Mathematical Statistics, 1940, 11, 271-283.

Olkin, I. Correlations revisited. In J.C. Stanley (Ed.), Improving experimental design and statistical analysis. Chicago: Rand McNally, 1967.

Spaeth, J.L. Path analysis. In D.J. Amick & H.J. Walberg (Eds.), Introductory multivariate analysis. Berkeley: McCutchan, 1975.

Williams, E.J. The comparison of regression variables. Journal of the Royal Statistical Society, 1959, 21, 396-399.

Equation (6)

$$\underline{t} = \frac{(\underline{r}_{i(j.k)} - \underline{r}_{i(j.m)})\left[(\underline{N} - 4)\left(1 + \frac{1 - \underline{r}_{jm}^{2} - \underline{r}_{jk}^{2} + \underline{r}_{jk}\underline{r}_{jm}\underline{r}_{km}}{(1 - \underline{r}_{jk}^{2})^{1/2}(1 - \underline{r}_{jm}^{2})^{1/2}}\right)\right]^{1/2}}{\left[2\left(1 - \underline{r}_{i(j.k)}^{2} - \underline{r}_{i(j.m)}^{2} - \frac{(1 - \underline{r}_{jm}^{2} - \underline{r}_{jk}^{2} + \underline{r}_{jk}\underline{r}_{km}\underline{r}_{jm})^{2}}{(1 - \underline{r}_{jk}^{2})(1 - \underline{r}_{jm}^{2})}\right) + 2\underline{r}_{i(j.k)}\underline{r}_{i(j.m)}\left(\frac{1 - \underline{r}_{jm}^{2} - \underline{r}_{jk}^{2} + \underline{r}_{jk}\underline{r}_{km}\underline{r}_{jm}}{(1 - \underline{r}_{jk}^{2})^{1/2}(1 - \underline{r}_{jm}^{2})^{1/2}}\right)\right]^{1/2}},$$

Appendix A

Let the following denote the nth individual's ($n = 1, 2, \ldots, N$) deviation scores on variables X_i, X_j, and X_k, respectively: $x_{in} = X_{in} - \bar{X}_i$, $x_{jn} = X_{jn} - \bar{X}_j$, and $x_{kn} = X_{kn} - \bar{X}_k$. Similarly, then, $x_{j*n} = x_{jn} - r_{jk}(\sigma_j/\sigma_k)x_{kn}$ and $x_{k*n} = x_{kn} - r_{jk}(\sigma_k/\sigma_j)x_{jn}$. Thus,

$$r_{j*k*} = \frac{\sum_{n=1}^{N}[x_{jn} - r_{jk}(\sigma_j/\sigma_k)x_{kn}][x_{kn} - r_{jk}(\sigma_k/\sigma_j)x_{jn}]}{(N-1)\sigma_{j*}\sigma_{k*}}$$

$$= \frac{\sum_{n=1}^{N} x_{jn}x_{kn} - r_{jk}(\sigma_j/\sigma_k)\sum_{n=1}^{N} x_{kn}^2 - r_{jk}(\sigma_k/\sigma_j)\sum_{n=1}^{N} x_{jn}^2 + r_{jk}^2\sum_{n=1}^{N} x_{jn}x_{kn}}{(N-1)\sigma_j\sigma_k(1 - r_{jk}^2)}$$

$$= (r_{jk} - r_{jk}^3)/(1 - r_{jk}^2).$$

Appendix B

Defining $\underline{x}_{j*n}$ as in Appendix A and $\underline{x}_{j**n} = \underline{x}_{jn} - \underline{r}_{jm}(\sigma_j/\sigma_m)\underline{x}_{mn}$,

$$\underline{r}_{j*j**} = \frac{\sum_{n=1}^{N} [\underline{x}_{jn} - \underline{r}_{jk}(\sigma_j/\sigma_k)\underline{x}_{kn}][\underline{x}_{jn} - \underline{r}_{jm}(\sigma_j/\sigma_m)\underline{x}_{mn}]}{(N-1)\sigma_{j*}\sigma_{j**}}$$

$$= \frac{\sum_{n=1}^{N}\underline{x}_{jn}^2 - \frac{\underline{r}_{jm}\sigma_j}{\sigma_m}\sum_{n=1}^{N}\underline{x}_{jn}\underline{x}_{mn} - \frac{\underline{r}_{jk}\sigma_j}{\sigma_k}\sum_{n=1}^{N}\underline{x}_{jn}\underline{x}_{kn} + \frac{\underline{r}_{jm}\underline{r}_{jk}\sigma_j^2}{\sigma_k\sigma_m}\sum_{n=1}^{N}\underline{x}_{kn}\underline{x}_{mn}}{(N-1)\sigma_j^2(1-\underline{r}_{jk}^2)^{1/2}(1-\underline{r}_{jm}^2)^{1/2}}$$

$$= \frac{1 - \underline{r}_{jm}^2 - \underline{r}_{jk}^2 + \underline{r}_{jk}\underline{r}_{km}\underline{r}_{jm}}{(1-\underline{r}_{jk}^2)^{1/2}(1-\underline{r}_{jm}^2)^{1/2}}.$$